

Parametrizing DVCS Form Factors with Neural Networks

Krešimir Kumerički

Physics Department
University of Zagreb, Croatia

Collaboration with:

Dieter Müller (Brookhaven National Laboratory),
Andreas Schäfer (Universität Regensburg)

28th September 2011, Ringberg Workshop "New Trends in HERA Physics 2011"

Outline

Introduction to Generalized Parton Distributions (GPDs) and Deeply Virtual Compton Scattering (DVCS)

Model-dependent global analysis of unpolarized target DVCS data

[Nucl. Phys. B841 (2010) 1-58, [arXiv:0904.0458](#)]

Neural networks approach [[JHEP 07 \(2011\) 073](#), [arXiv:1106.2808](#)]

Probing the proton with two photons

- Deeply virtual Compton scattering (DVCS) [Müller '92, et al. '94]

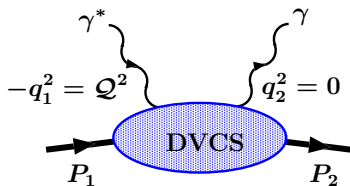
$$P = P_1 + P_2, \quad t = (P_2 - P_1)^2$$

$$q = (q_1 + q_2)/2$$

Generalized Bjorken limit:

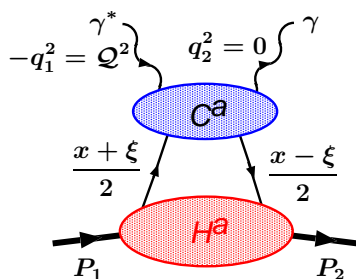
$$-q^2 \simeq Q^2/2 \rightarrow \infty$$

$$\xi = \frac{-q^2}{2P \cdot q} \rightarrow \text{const}$$



- To leading twist-two accuracy cross-section can be expressed in terms of **Compton form factors** (CFFs)

$$\mathcal{H}(\xi, t, Q^2), \mathcal{E}(\xi, t, Q^2), \tilde{\mathcal{H}}(\xi, t, Q^2), \tilde{\mathcal{E}}(\xi, t, Q^2), \dots$$

Factorization of DVCS \longrightarrow GPDs

$$P = P_1 + P_2, \quad t = (P_2 - P_1)^2$$

$$q = (q_1 + q_2)/2$$

$$-q^2 \simeq Q^2/2 \rightarrow \infty$$

$$\xi = \frac{-q^2}{2P \cdot q} \rightarrow \text{const}$$

- Compton form factor is a convolution:

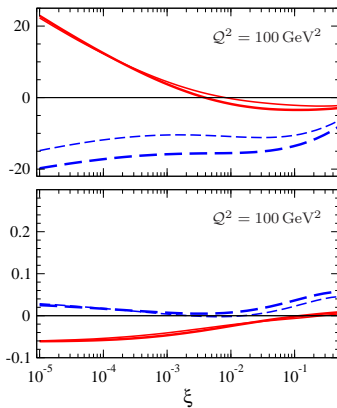
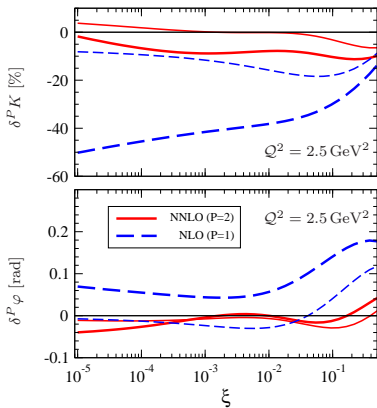
$$\mathcal{H}(\xi, t, Q^2) = \int dx C^a(x, \xi, Q^2/Q_0^2) H^a(x, \xi, t, Q_0^2)$$

$$a = \text{NS}, S(\Sigma, G)$$

- $H^a(x, \xi, t, Q_0^2)$ — Generalized parton distribution (GPD)

(N)NLO corrections

[K.K., D. Müller, K. Passek-K. '07]



Thick lines:
 "hard" gluon
 $N_G = 0.4$
 $\alpha_G(0) = \alpha_\Sigma(0) + 0.05$

Thin lines:
 "soft" gluon
 $N_G = 0.3$
 $\alpha_G(0) = \alpha_\Sigma(0) - 0.02$

$$\delta^P K = \frac{|\mathcal{H}^{N^P \text{LO}}|}{|\mathcal{H}^{N^{P-1} \text{LO}}|} - 1, \quad \delta^P \varphi = \arg\left(\frac{\mathcal{H}^{N^P \text{LO}}}{\mathcal{H}^{N^{P-1} \text{LO}}}\right)$$

Dispersion-relation access to GPDs at LO

[Teryaev '05; K.K., Müller and Passek-K. '07, '08; Diehl and Ivanov '07]

- LO perturbative prediction is “handbag” amplitude

$$\mathcal{H}(\xi, t, Q^2) \stackrel{\text{LO}}{=} \int_{-1}^1 dx \left(\frac{1}{\xi - x - i\epsilon} - \frac{1}{\xi + x - i\epsilon} \right) H(x, \xi, t, Q^2)$$

- giving access to GPD on the “cross-over” line $\xi = x$

$$\frac{1}{\pi} \Im \mathcal{H}(\xi, t, Q^2) \stackrel{\text{LO}}{=} H(\xi, \xi, t, Q^2) - H(-\xi, \xi, t, Q^2)$$

- while dispersion relation connects it to $\Re \mathcal{H}$ and at the most one subtraction constant $\mathcal{C}_{\mathcal{H}} = -\mathcal{C}_{\mathcal{E}}$; $\mathcal{C}_{\tilde{\mathcal{H}}} = \mathcal{C}_{\tilde{\mathcal{E}}} = 0$

$$\begin{aligned} \Re \mathcal{H}(\xi, t, Q^2) = \\ \frac{1}{\pi} \text{PV} \int_0^1 d\xi' \left(\frac{1}{\xi - \xi'} - \frac{1}{\xi + \xi'} \right) \Im \mathcal{H}(\xi', t, Q^2) + \mathcal{C}_{\mathcal{H}}(t, Q^2) \end{aligned}$$

Model-dependent extraction of GPDs

- Revealing GPD H from DVCS on *unpolarized* proton target at LO [K.K. and D. Müller '09]

$$\Im \mathcal{H}(\xi, t) = \pi \left[\left(2\frac{4}{9} + \frac{1}{9} \right) H^{\text{val}}(\xi, \xi, t) + \frac{2}{9} H^{\text{sea}}(\xi, \xi, t) \right]$$

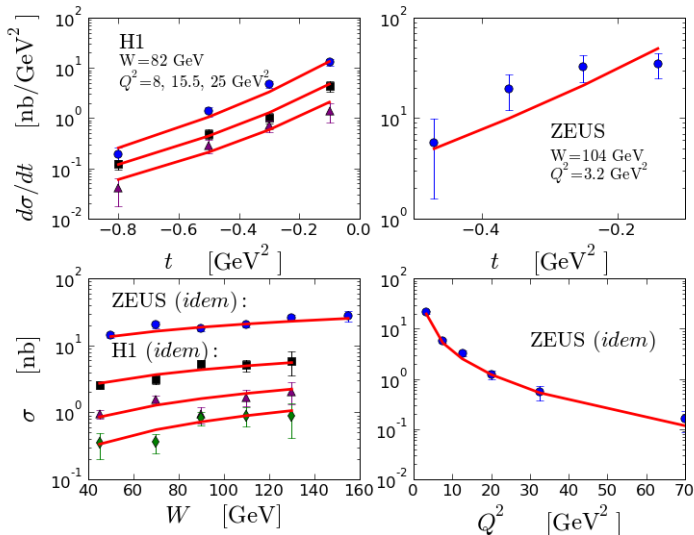
- Valence** quarks model (ignoring Q^2 evolution):

$$H^{\text{val}}(\xi, \xi, t) = n \left(\frac{2\xi}{1+\xi} \right)^{-\alpha^{\text{val}}(t)} \left(\frac{1-\xi}{1+\xi} \right)^b \frac{1}{\left(1 - \frac{1-\xi}{1+\xi} \frac{t}{M^2} \right)}$$

$$\alpha^{\text{val}}(t) = 0.43 + 0.85 t/\text{GeV}^2 \quad (\rho, \omega) \text{ Regge}$$

- Sea** partons modelled similarly, but in moment space + t -channel partial wave expansion + LO Q^2 evolution
- 8–13 parameter**-fit: $\chi^2/d.o.f. = 132/160$

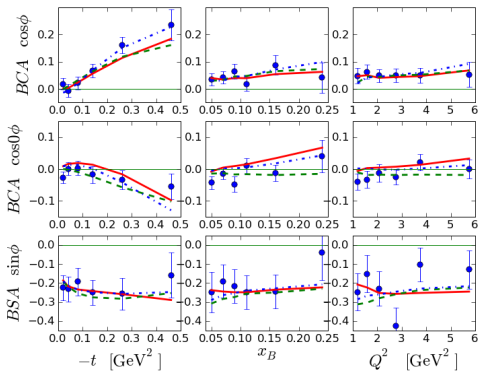
H1 (2007), ZEUS (2008)



HERMES (2008)

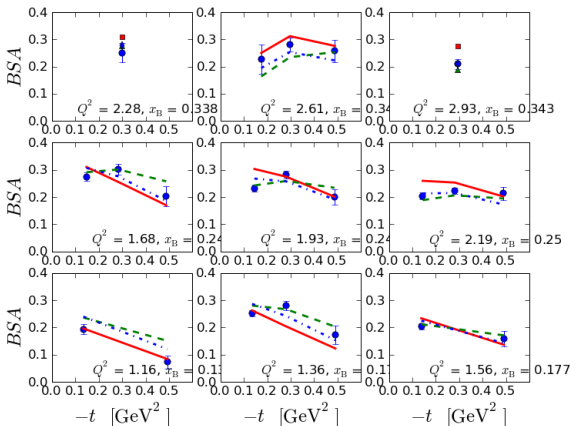
$$BCA \equiv \frac{d\sigma_{e^+} - d\sigma_{e^-}}{d\sigma_{e^+} + d\sigma_{e^-}} \sim A_C^{\cos 0\phi} + A_C^{\cos 1\phi} \cos \phi \sim \text{Re}\mathcal{H}$$

$$BSA \equiv \frac{d\sigma_{e^\uparrow} - d\sigma_{e^\downarrow}}{d\sigma_{e^\uparrow} + d\sigma_{e^\downarrow}} \sim A_{LU}^{\sin 1\phi} \sin \phi \sim \text{Im}\mathcal{H}$$



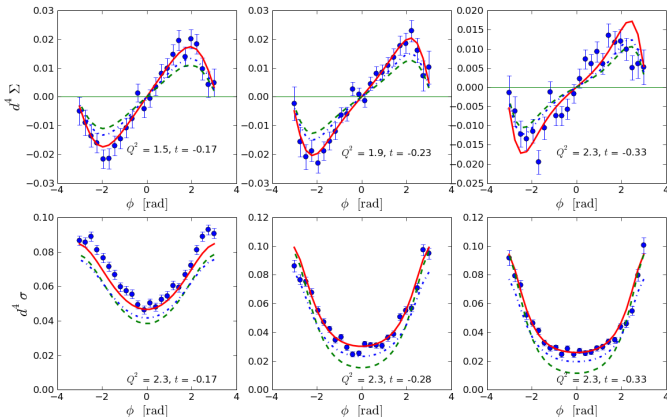
CLAS (2007)

- BSA. (Only data with $|t| \leq 0.3 \text{ GeV}^2$ used for fits.)

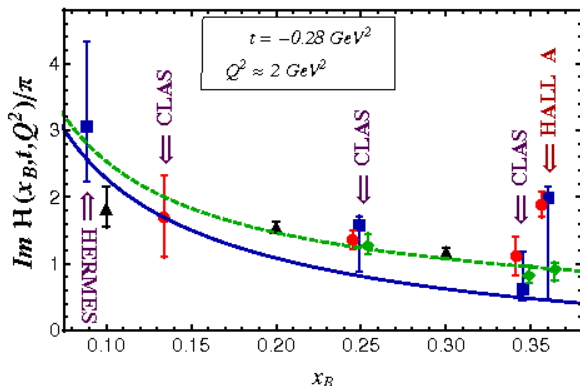


Hall A (2006)

- Fit to **unpolarized cross section** $d\sigma/(dx_B dQ^2 dt d\phi) \sim \Re\mathcal{H}$
- Fit is OK only with unusually large $\Re\mathcal{T}_{\text{DVCS}} (\rightarrow \tilde{\mathcal{H}})$



Result and comparison to others



[Guidal '08, Guidal and Moutarde '09], seven CFF fit (blue squares), [Guidal '10] $\mathcal{H}, \tilde{\mathcal{H}}$ CFF fit (green diamonds), [Moutarde '09] H GPD fit (red circles)

Models are available at WWW

- <http://calculon.phy.hr/gpd/>

```
% xs.exe
```

```
xs.exe ModelID Charge Polarization Ee Ep xB Q2 t phi
```

returns cross section (in nb) for scattering of lepton of energy Ee on unpolarized proton of energy Ep. Charge=-1 is for electron.

ModelID is one of

- 0 debug, always returns 42,
- 1 KM09a - arXiv:0904.0458 fit without Hall A,
- 2 KM09b - arXiv:0904.0458 fit with Hall A,
- 3 KM10 - preliminary hybrid fit with LO sea evolution, from Trento presentation,
- 4 KM10a - preliminary hybrid fit with LO sea evolution, without Hall A data
- 5 KM10b - preliminary hybrid fit with LO sea evolution, with Hall A data

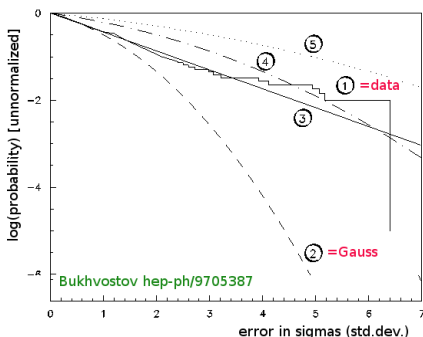
xB Q2 t phi -- usual kinematics (phi is in Trento convention)

```
% xs.exe 1 -1 1 27.6 0.938 0.111 3. -0.3 0
```

```
0.18584386497251
```

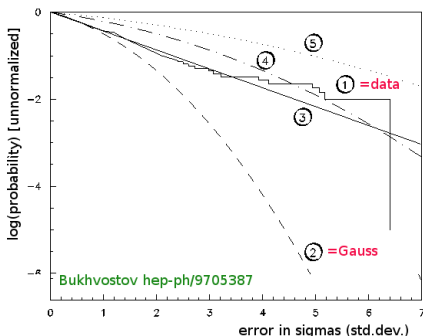
Problems with standard fitting approaches

1. Choice of fitting function introduces **theoretical bias** leading to **systematic error** which cannot be estimated (and is likely much larger for GPDs(x, ξ, t) than for PDFs(x)).
2. **Propagation of uncertainties** from experiment to fitted function is difficult. Errors in actual experiments are not always Gaussian.

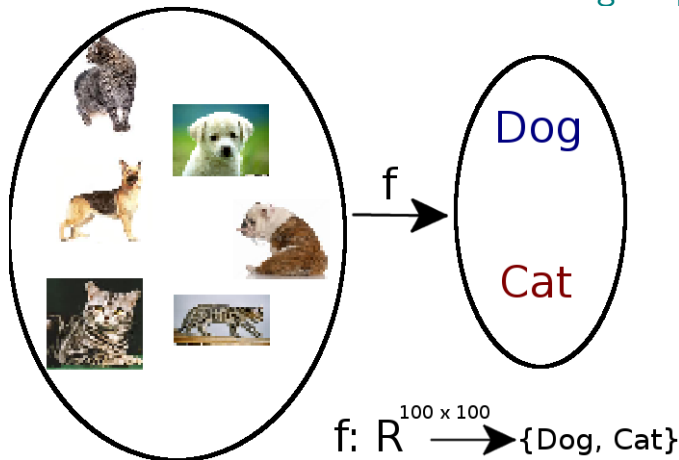


Problems with standard fitting approaches

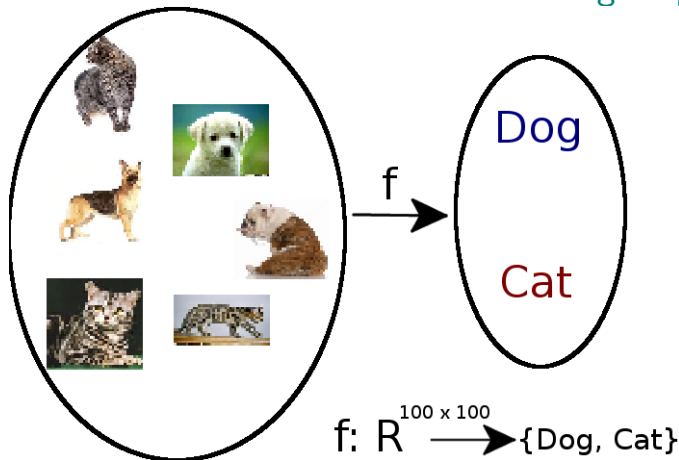
1. Choice of fitting function introduces theoretical bias leading to systematic error which cannot be estimated (and is likely much larger for GPDs(x, ξ, t) than for PDFs(x)). → **NNets**
2. Propagation of uncertainties from experiment to fitted function is difficult. Errors in actual experiments are not always Gaussian. → **Monte Carlo error propagation**



Introduction to neural networks: Cat-or-dog mapping

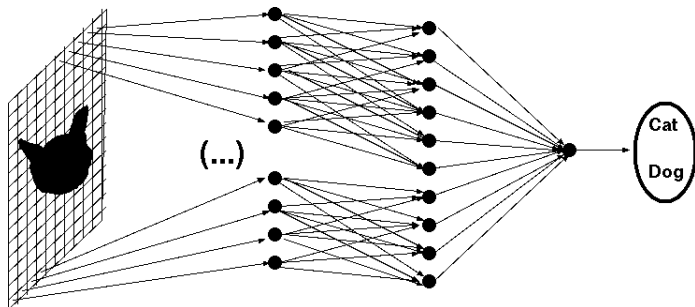


Introduction to neural networks: Cat-or-dog mapping



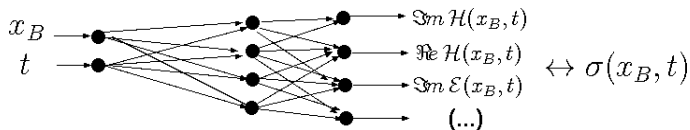
- How to represent function f by a computer algorithm?
- \longrightarrow neural networks, learning machines, AI

Cat-or-dog mapping by neural network



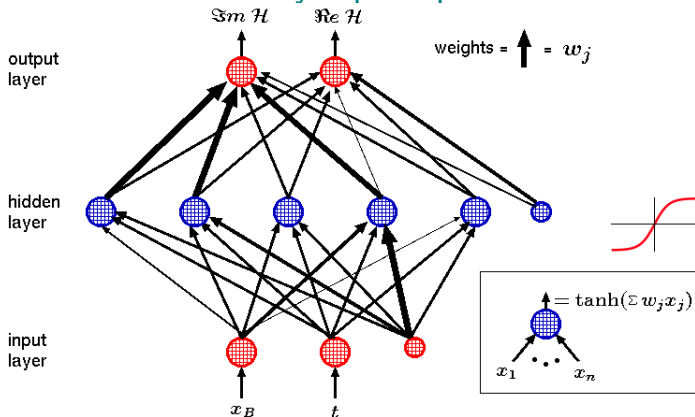
- Parameters (“weights”) of neural network adjusted by “training” it on many samples
- Neural network becomes a representation of function f .
- Neural networks are capable of generalization: they successfully classify objects not seen during training

Neural networks as a fitting tool

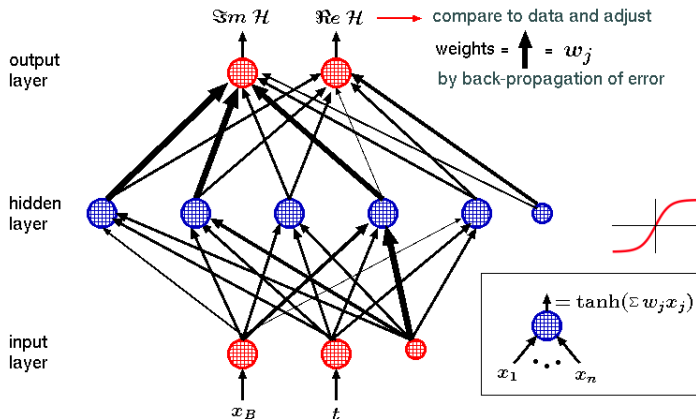


- Neural network now represents mapping $f : \mathbb{R}^2 \rightarrow \mathbb{R}^{n_{\mathcal{F}}}$.
- Classification problem is just a special case of optimization (χ^2 minimization) problem (where we have $\sigma(x_B, t) \in \mathbb{R}$ instead of *output* $\in \{\text{cat}, \text{dog}\}$).
- We can hope to be able to train neural networks to represent real underlying physical law
- NN approach is successfully applied to PDF fitting by [NNPDF] group and should be even more powerful in GPD fitting with GPDs being less-known functions of **more variables**.

Multilayer perceptron



Multilayer perceptron



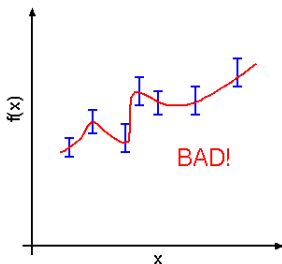
- Essentially a least-square fit of a complicated many-parameter function. $f(x) = \tanh(\sum w_i \tanh(\sum w_j \dots)) \Rightarrow$ no theory bias

Function fitting by a neural net

- **Theorem:** Given enough neurons, any smooth function $f(x_1, x_2, \dots)$ can be approximated to any desired accuracy. Single hidden layer is sufficient (but not always most efficient).

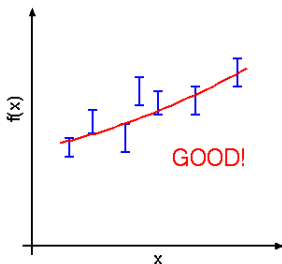
Function fitting by a neural net

- **Theorem:** Given enough neurons, any smooth function $f(x_1, x_2, \dots)$ can be approximated to any desired accuracy. Single hidden layer is sufficient (but not always most efficient).
- With simple training of neural nets to data there is a danger of **overfitting** (a.k.a. overtraining)



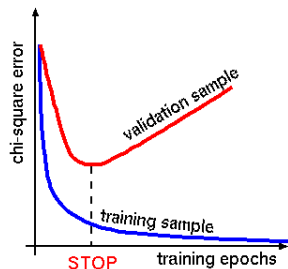
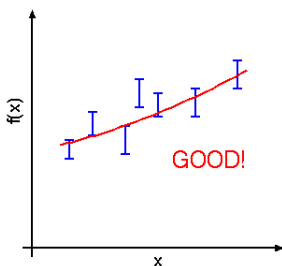
Function fitting by a neural net

- **Theorem:** Given enough neurons, any smooth function $f(x_1, x_2, \dots)$ can be approximated to any desired accuracy. Single hidden layer is sufficient (but not always most efficient).
- With simple training of neural nets to data there is a danger of **overfitting** (a.k.a. overtraining)

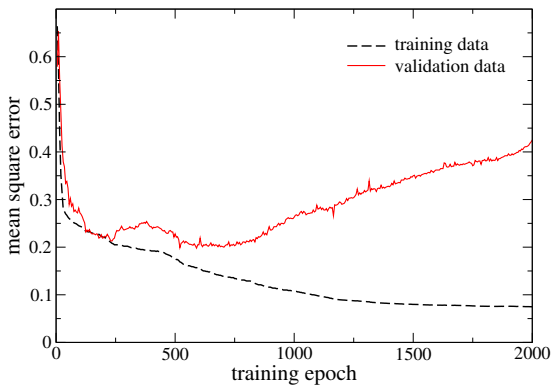


Function fitting by a neural net

- **Theorem:** Given enough neurons, any smooth function $f(x_1, x_2, \dots)$ can be approximated to any desired accuracy. Single hidden layer is sufficient (but not always most efficient).
- With simple training of neural nets to data there is a danger of **overfitting** (a.k.a. overtraining)
- **Solution:** Divide data (randomly) into two sets: *training* sample and *validation* sample. Stop training when error of validation sample starts increasing.



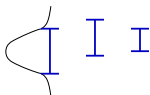
Example of a training with crossvalidation



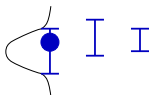
Monte Carlo propagation of errors

I I I

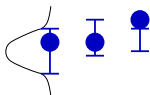
Monte Carlo propagation of errors



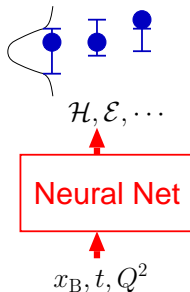
Monte Carlo propagation of errors



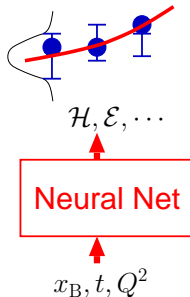
Monte Carlo propagation of errors



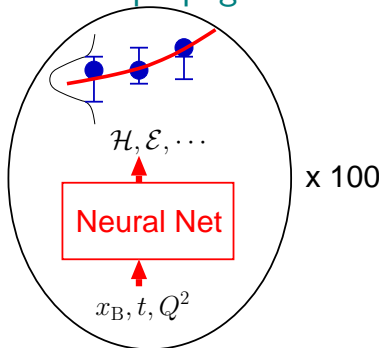
Monte Carlo propagation of errors



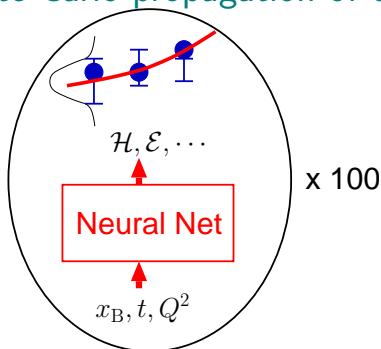
Monte Carlo propagation of errors



Monte Carlo propagation of errors



Monte Carlo propagation of errors



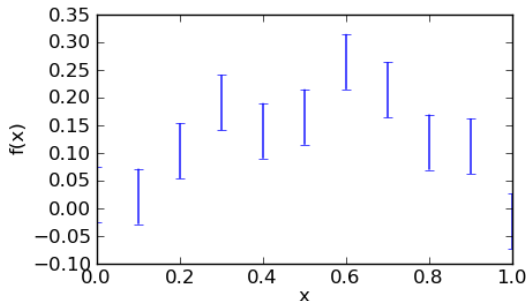
- Set of N_{rep} NNs defines a probability distribution in space of possible CFF functions:

$$\langle \mathcal{F}[\mathcal{H}] \rangle = \int \mathcal{D}\mathcal{H} \mathcal{P}[\mathcal{H}] \mathcal{F}[\mathcal{H}] = \frac{1}{N_{rep}} \sum_{k=1}^{N_{rep}} \mathcal{F}[\mathcal{H}^{(k)}], \quad (1)$$

- Experimental uncertainties and their correlations are preserved [Giele et al., '01]

Fitting example

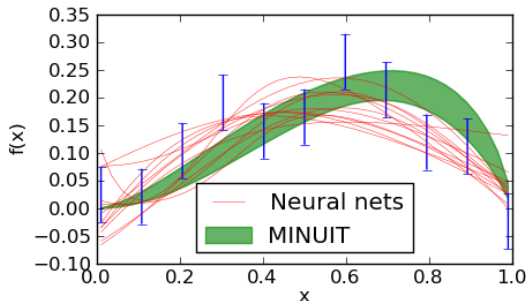
- Take some (fake) data



- Fit with
 1. Standard Minuit fit with ansatz $f(x) = x^a(1-x)^b$
 2. Neural network fit

Fitting example

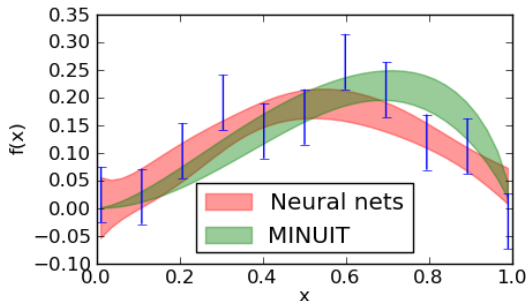
- Take some (fake) data



- Fit with
 - Standard Minuit fit with ansatz $f(x) = x^a(1-x)^b$
 - Neural network fit

Fitting example

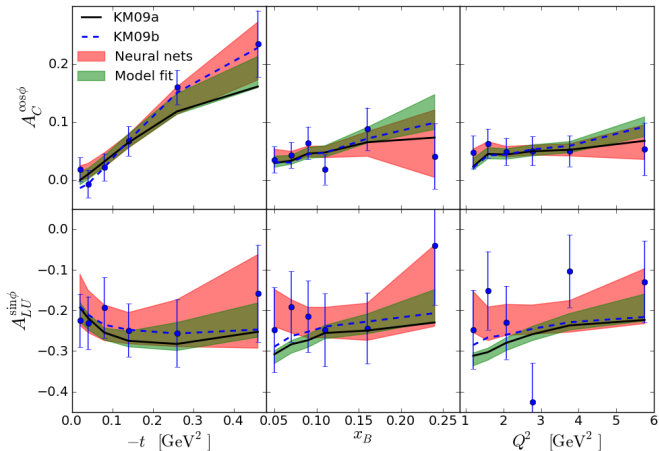
- Take some (fake) data



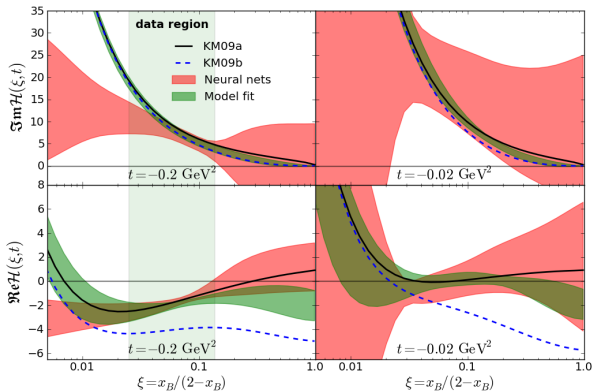
- Fit with
 - Standard Minuit fit with ansatz $f(x) = x^a(1-x)^b$
 - Neural network fit

Fit to actual HERMES BSA+BCA data

- 50 neural nets with 13 neurons in a single hidden layer

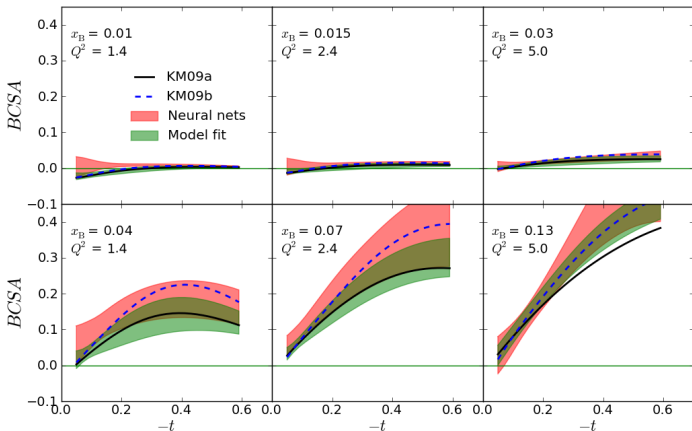


Resulting neural network CFFs



Prediction for COMPASS II BCSA

$$BCSA = \frac{d\sigma_{\mu\downarrow+} - d\sigma_{\mu\uparrow-}}{d\sigma_{\mu\downarrow+} + d\sigma_{\mu\uparrow-}} \quad (E_{\mu} = 160 \text{ GeV})$$



Curse of dimensionality

- *It is relatively easy to find a coin lying somewhere on 100 meter string. It is very difficult to find it on a football field.*

Curse of dimensionality

- *It is relatively easy to find a coin lying somewhere on 100 meter string. It is very difficult to find it on a football field.*
- Similarly, in contrast to $PDFs(x)$, it is very difficult to perform truly model independent extraction of $CFFs(\xi, t)$ or $GPDs(x, \xi, t)$.
- Known GPD constraints don't decrease the dimensionality of the GPD domain space.
- Model fitting cannot be discarded. Both approaches have their advantages.
- Neural net results can guide us in model building.

Summary

- Global dispersion-relation fits of all unpolarized proton DVCS data are possible with assumption of GPD H dominance; only unpolarized-beam cross sections measured by Hall A require some additional contributions (like \tilde{H}).
- Neural networks offer a powerful alternative approach to extraction of hadron structure information from measurements, enabling model-independent fits and facilitating error propagation from data to resulting CFFs/GPDs.

Summary

- Global dispersion-relation fits of all unpolarized proton DVCS data are possible with assumption of GPD H dominance; only unpolarized-beam cross sections measured by Hall A require some additional contributions (like \tilde{H}).
- Neural networks offer a powerful alternative approach to extraction of hadron structure information from measurements, enabling model-independent fits and facilitating error propagation from data to resulting CFFs/GPDs.

The End

Some properties of GPDs

- Forward limit ($\Delta \rightarrow 0$): \Rightarrow GPD \rightarrow PDF

$$F^q(x, 0, 0) = H^q(x, 0, 0) = \theta(x)q(x) - \theta(-x)\bar{q}(-x)$$

- Polynomiality:

$$\int_{-1}^1 dx x^j H^q(x, \eta, t) = \sum_{k=0, \text{even}}^j (2\eta)^k A_{j+1, k}^q(t) \quad (\text{even } j)$$

- Sum rules:

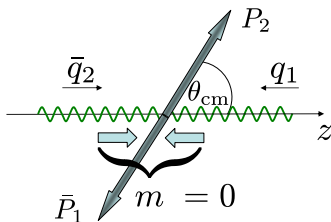
$$\int_{-1}^1 dx \begin{cases} H^q(x, \eta, t) \\ E^q(x, \eta, t) \end{cases} = \begin{cases} F_1^q(t) & \text{Dirac} \\ F_2^q(t) & \text{Pauli} \end{cases}$$

- “Ji’s sum rule” (related to proton spin problem)

$$J^q = \frac{1}{2} \int_{-1}^1 dx x \left[H^q(x, \eta, t) + E^q(x, \eta, t) \right]_{t \rightarrow 0} \quad [\text{Ji '96}]$$

Modelling conformal moments of GPDs (I)

- How to model η -dependence of GPD's $H_j(\eta, t)$?
- Idea: consider crossed t -channel process $\gamma^* \gamma \rightarrow p \bar{p}$

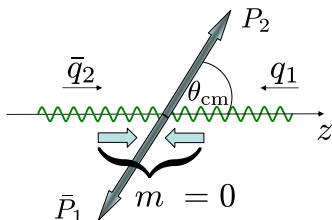


When crossing back to DVCS channel we have:

$$\cos \theta_{\text{cm}} \rightarrow -\frac{1}{\eta}$$

Modelling conformal moments of GPDs (I)

- How to model η -dependence of GPD's $H_j(\eta, t)$?
- Idea: consider crossed t -channel process $\gamma^* \gamma \rightarrow p \bar{p}$



When crossing back to DVCS channel we have:

$$\cos \theta_{\text{cm}} \rightarrow -\frac{1}{\eta}$$

- ... and dependence on θ_{cm} in t -channel is given by $\text{SO}(3)$ partial wave decomposition of $\gamma^* \gamma$ scattering

$$\mathcal{H}(\eta, \dots) = \mathcal{H}^{(t)}(\cos \theta_{\text{cm}} = -\frac{1}{\eta}, \dots) = \sum_J (2J+1) f_J(\dots) d_{0,\nu}^J(\cos \theta)$$

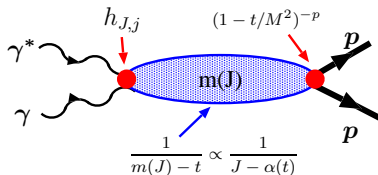
- $d_{0,\nu}^J$ — Wigner $\text{SO}(3)$ functions (Legendre, Gegenbauer, ...)
- $\nu = 0, \pm 1$ — depending on hadron helicities

Modelling conformal moments of GPDs (II)

- OPE expansion of both \mathcal{H} and $\mathcal{H}^{(t)}$ leads to

$$H_j(\eta, t) = \eta^{j+1} H_j^{(t)}(\cos\theta = -\frac{1}{\eta}, s^{(t)} = t)$$

- and t -channel partial waves are modelled as:



$$H_j(\eta, t) = \sum_J^{j+1} h_{J,j} \frac{1}{J - \alpha(t)} \frac{1}{\left(1 - \frac{t}{M^2(J)}\right)^p} \eta^{j+1-J} d_{0,\nu}^J$$

- Similar to “dual” parametrization [Polyakov, Shuvaev '02]

I-PW model — only leading partial wave

- Taking just a leading partial wave $J = j + 1$ gives ansatz:

$$\mathbf{H}_j(\xi, t, \mu_0^2) = \begin{pmatrix} N'_\Sigma F_\Sigma(t) \text{B}(1 + j - \alpha_\Sigma(0), 8) \\ N'_G F_G(t) \text{B}(1 + j - \alpha_G(0), 6) \end{pmatrix}$$

$$\alpha_a(t) = \alpha_a(0) + 0.15t \quad F_a(t) = \frac{j + 1 - \alpha(0)}{j + 1 - \alpha(t)} \left(1 - \frac{t}{M_0^{a2}}\right)^{-p_a}$$

... corresponding in forward case to **PDFs** of form

$$\Sigma(x) = N'_\Sigma x^{-\alpha_\Sigma(0)} (1 - x)^7; \quad G(x) = N'_G x^{-\alpha_G(0)} (1 - x)^5$$

I-PW model — only leading partial wave

- Taking just a leading partial wave $J = j + 1$ gives ansatz:

$$\mathbf{H}_j(\xi, t, \mu_0^2) = \begin{pmatrix} N'_\Sigma F_\Sigma(t) B(1+j-\alpha_\Sigma(0), 8) \\ N'_G F_G(t) B(1+j-\alpha_G(0), 6) \end{pmatrix}$$

$$\alpha_a(t) = \alpha_a(0) + 0.15t \quad F_a(t) = \frac{j+1-\alpha(0)}{j+1-\alpha(t)} \left(1 - \frac{t}{M_0^{a2}}\right)^{-p_a}$$

... corresponding in forward case to PDFs of form

$$\Sigma(x) = N'_\Sigma x^{-\alpha_\Sigma(0)} (1-x)^7; \quad G(x) = N'_G x^{-\alpha_G(0)} (1-x)^5$$

- $M_0^G = \sqrt{0.7} \text{ GeV}$ is fixed by the J/ψ production data
- Free parameters: N'_Σ , $\alpha_\Sigma(0)$, M_0^Σ , N'_G , $\alpha_G(0)$

For small ξ (small x_{Bj}) valence quarks are less important $\Rightarrow \Sigma \approx \text{sea}$

Inclusion of subleading PW — flexible models

- [K.K. and D. Müller '09]

$$\mathbf{H}_j(\eta, t) = \underbrace{\begin{pmatrix} N'_{\text{sea}} F_{\text{sea}}(t) B(1+j-\alpha_{\text{sea}}(0), 8) \\ N'_G F_G(t) B(1+j-\alpha_G(0), 6) \end{pmatrix}}_{\text{skewness } r \approx 1.6 \text{ (too large)}} + \underbrace{\begin{pmatrix} s_{\text{sea}} \\ s_G \end{pmatrix}}_{\substack{< 0 \\ \text{negative skewness}}} \left(\begin{array}{l} \text{subleading par-} \\ \text{tial waves, } \eta\text{-} \\ \text{dependence!} \end{array} \right)$$

- nl-PW — addition of second PW needed for good fits
- two new parameters: s_{sea} and s_G
- nnl-PW — addition of third PW (doesn't improve fits but makes possible positive gluon GPDs at small Q^2).

Beam charge asymmetry

$$BCA \equiv \frac{d\sigma_{e^+} - d\sigma_{e^-}}{d\sigma_{e^+} + d\sigma_{e^-}} = \frac{\mathcal{A}_{\text{Interference}}}{|\mathcal{A}_{\text{DVCS}}|^2 + |\mathcal{A}_{\text{BH}}|^2} \stackrel{\text{LO}}{\propto} F_1 \Re \mathcal{H} + \frac{|t|}{4M^2} F_2 \Re \mathcal{E}$$

Beam charge asymmetry

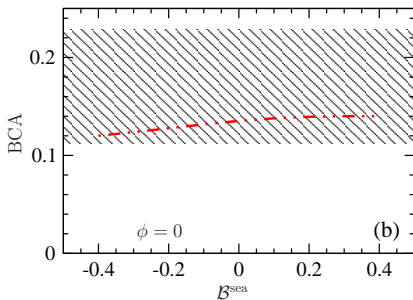
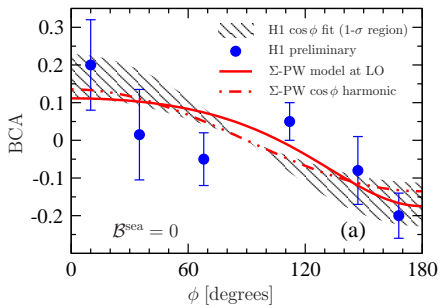
$$BCA \equiv \frac{d\sigma_{e^+} - d\sigma_{e^-}}{d\sigma_{e^+} + d\sigma_{e^-}} = \frac{\mathcal{A}_{\text{Interference}}}{|\mathcal{A}_{\text{DVCS}}|^2 + |\mathcal{A}_{\text{BH}}|^2} \stackrel{\text{LO}}{\propto} F_1 \text{Re}\mathcal{H} + \frac{|t|}{4M^2} F_2 \text{Re}\mathcal{E}$$

- Model E_{sea} as $(\mathcal{B}_{\text{sea}}/N_{\text{sea}})H_{\text{sea}}$ and take $\mathcal{B}_{\text{sea}} \equiv \int dx x E_{\text{sea}}$ as a parameter

Beam charge asymmetry

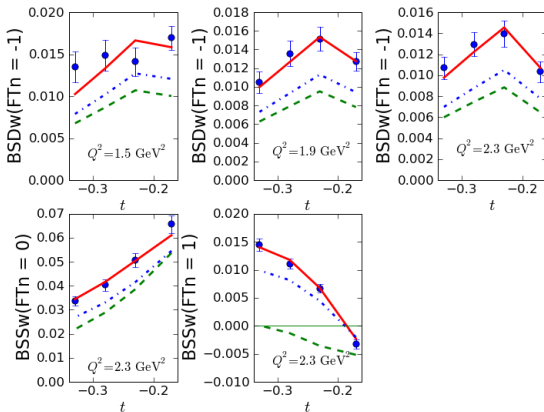
$$BCA \equiv \frac{d\sigma_{e^+} - d\sigma_{e^-}}{d\sigma_{e^+} + d\sigma_{e^-}} = \frac{\mathcal{A}_{\text{Interference}}}{|\mathcal{A}_{\text{DVCS}}|^2 + |\mathcal{A}_{\text{BH}}|^2} \stackrel{\text{LO}}{\propto} F_1 \text{Re}\mathcal{H} + \frac{|t|}{4M^2} F_2 \text{Re}\mathcal{E}$$

- Model E_{sea} as $(\mathcal{B}_{\text{sea}}/N_{\text{sea}})H_{\text{sea}}$ and take $\mathcal{B}_{\text{sea}} \equiv \int dx x E_{\text{sea}}$ as a parameter



- We cannot extract \mathcal{B}_{sea} from H1 data

Hall A (2006) II



Prediction for EIC cross section

